Exchangeable Urn Processes*

by

Bruce M. Hill, David Lane and William Sudderth
University of Michigan and University of Minnesota

Technical Report No. 452
July 1985

*Research supported by National Science Foundation Grant MCS-8300585 (for Hill) and DMS-8421208 (for Sudderth).

Abstract

If Y_n is 1 or 0 depending on whether the nth ball drawn in a Polya urn scheme is red or not, then the variables $Y_1, Y_2, \ldots$ are exchangeable. It is shown for a generalized class of urn models that no other scheme gives rise to exchangeable variables unless the Y_n are either independent and identically distributed, or deterministic (that is, all of the Y_n's have the same value with probability 1).

AMS 1980 subject classification. Primary 60G09, 62A15.

Key words and phrases. Exchangeable, generalized urn process, Polya urn.

1. Introduction.

Let $Y = (Y_1, Y_2, \ldots)$ be a sequence of $\{0,1\}$-valued random variables. The process Y is exchangeable if its distribution is invariant under finite permutations of the indices. The notion of exchangeability was introduced by de Finetti (1931, 1937) and is fundamental to subjective probability and Bayesian statistics. As de Finetti (1931) showed, for every exchangeable process Y, there is a random variable Θ with values in $[0,1]$ such that, given $\Theta = \theta$, the variables $Y_1, Y_2, \ldots$ are independent Bernoulli (θ). (A convenient reference is section 2 of Freedman (1965).) The distribution Q of the variable Θ will be called the de Finetti measure for the exchangeable process Y.

Conversely, an exchangeable process can be constructed according to the following recipe: suppose Θ is a random variable with values in $[0,1]$ and let $Y_1, Y_2, \ldots$ be conditionally independent Bernoulli (θ) variables given $\Theta = \theta$. Then the process Y is obviously exchangeable.

The Polya urn scheme is another interesting way to generate a sequence of $\{0,1\}$-valued random variables. Suppose that an urn initially contains r red and b black balls and that, at each stage, a ball is selected at random and replaced by two of the same color. Let Y_n be 1 or 0 accordingly as the nth ball selected is red or black. Then the Polya process Y is easily seen to be exchangeable and has a de Finetti measure which is a beta distribution with parameters r and b (see Polya (1931), or Freedman (1965)). Thus, two different methods for generating data, corresponding to the Polya urn scheme and the scheme discussed in the previous paragraph (with a beta distribution for Θ) give rise to precisely the same distribution for the process Y. (For an interesting discussion of the connection between these two schemes, see de Finetti (1975),

p. 220).

The Polya process is a special case of a family of processes whose distinguishing feature is that the sequence of observations can be concretely represented by sucessive drawings from urns of changing compositions. To define this family precisely, consider an urn with initial composition (r,b) of r red balls and b black balls and let f be a mapping from the unit interval to itself. Set $X_0 = r/(r+b)$, the initial proportion of red balls, and suppose that a red ball is added to the urn with probability $f(X_0)$ and a black ball is added with probability $1 - f(X_0)$. Let X_1 be the new proportion of red balls and iterate the procedure to generate a process $X = (X_0,X_1,X_2,\ldots)$. As before, let Y_n be the indicator of the event that the nth ball added is red. The process $Y = (Y_1,Y_2,\ldots)$ is an urn process with initial composition (r,b) and urn function f. The distribution of an urn process Y is completely determined by the initial composition and the values of the urn function f at the successive proportions $X_0,X_1,\ldots$. (The only possible proportions are of the form $(r+k)/(r+b+n)$ where k and n are non-negative integers and $k \leq n$.) These generalized urn schemes were introduced by Hill, Lane, and Sudderth (1980), who proved a convergence theorem for the process X. Generalized urns with balls of many colors have been studied by Arthur, Ermoliev, and Kaniovski (1983). See also Johnson and Kotz (1977) for a discussion of a variety of urn processes.

Notice that the process Y for the Polya urn scheme is an urn process with urn function $f(x) = x$. As we saw, such processes are exchangeable, with beta de Finetti measures. A constant urn function $f(x) \equiv p$ generates a Bernoulli process $Y_1,Y_2,\ldots$ of independent, Bernoulli (p) variables. Such a process is clearly exchangeable and has a de Finetti measure concentrated at the single point p. A trivial collection of exchangeable urn processes are the <u>deterministic</u> ones. Suppose initially a red ball is added with probability p and a black with probability $1 - p$, and that all subsequent balls are the same color as the first. This scheme corresponds to an urn function f which equals p at X_0, is identically 1 on $(X_0,1]$ and identically 0 on $[0,X_0)$. The de Finetti measure Q assigns probability p to $\{1\}$ and $1 - p$ to $\{0\}$.

On the other hand, not all urn processes are exchangeable. For example, if $f(x) = 1 - x$ and the initial urn composition is $(1,2)$, a simple calculation shows the probability of a one followed by a zero differs from the probability of a zero followed by a one. Hence, this urn process is not exchangeable.

Similarly, not all exchangeable processes can be represented as urn processes. For example, it is easy to see that a distribution for θ placing probability 1/2 on $\theta = 1/3$ and probability 1/2 on $\theta = 2/3$ can not be an urn process.

This paper addresses the question: which urn processes are exchangeable? The following theorem provides the answer.

<u>Theorem</u>. Suppose the process $Y = (Y_1,Y_2,\ldots)$ of $\{0,1\}$-valued random variables is an exchangeable urn process. Then Y is Polya, Bernoulli, or deterministic.

The proof is in the next section.

The result of the theorem can be expressed in a manner reminiscent of W.E. Johnson's Sufficientness Postulate (cf. Zabell (1982)). Johnson's Postulate assumes a finite sequence $Y_1,\ldots, Y_n$ of $\{0,1,\ldots, k-1\}$-valued random variables which are conditionally independent given the probability vector $p = (p_0, \ldots, p_{k-1})$. The postulate then states that, for $0 \le j \le k-1$,

$$P[Y_n = j \mid Y_1 = y_1, \ldots, Y_{n-1} = y_{n-1}]$$

depends only on n_j, the number of the y_i's which equal j. Johnson argued that the only prior distributions on p consistent with his postulate are symmetric Dirichlet distributions and point masses. This result is of course vacuous for $k = 2$.

Suppose however that $k = 2$ and the sequence can be infinitely extended. Assume further that, for every n,

$$P[Y_{n+1} = 1 \mid Y_1 = y_1, \ldots Y_n = y_n] \tag{1.1}$$

is a fixed function of x_n, the proportion of red balls in the urn. In this context, our Theorem shows that the only distributions for p consistent with this assumption are beta distributions, point masses, and distributions concentrated on $\{0,1\}$.

Let $S_n = Y_1 + \ldots + Y_n$. The proportion X_n of red balls at stage n can be written as

$$X_n = (r+S_n)/(r+b+n).$$

For urn processes, the predictive probability (1.1) is a function of X_n rather than S_n/n. Are there exchangeable processes Y for which the probability in

(1.1) is a fixed function of S_n/n? It follows easily from the theorem that the only such processes are Bernoulli or deterministic. (To see this, observe that given $Y_1 = 1$ (0), $Y_2, Y_3, \ldots$ is an exchangeable urn process with initial composition (1,0) ((0,1)). Apply the theorem and argue that the processes cannot be Polya.)

2. Proof of the theorem.

Throughout this section, $Y = (Y_1, Y_2, \ldots)$ is an exchangeable process of 0's and 1's with de Finetti measure Q. It is assumed that $Q\{0,1\} < 1$ and that Q is not concentrated at a single point. In the language of the previous section, we are assuming Y is neither Bernoulli nor deterministic.

Let $S_n = Y_1 + \ldots + Y_n$ for $n \geq 1$. Because of the first assumption about Q, all of the events $[S_n = k]$ have positive probability for $k = 0, 1, \ldots, n$ and $n \geq 1$.

Lemma 1. For $n \geq 1$, the function

$$g_n(k) = P[Y_{n+1}=1 | S_n=k]$$

is strictly increasing on its domain $\{0, 1, \ldots, n\}$.

Proof. Consider first the case when $n = 1$. Let Θ be a random variable with distribution Q. Then

$$P[Y_2=1|Y_1=1] - P[Y_2=1|Y_1=0] = \frac{E(\theta^2)}{E(\theta)} - \frac{E(\theta(1-\theta))}{E(1-\theta)}$$
$$= \frac{E(\theta^2) - (E\theta)^2}{E(\theta)E(1-\theta)}$$
$$> 0.$$

Next suppose $n \geq 2$ and $0 \leq k \leq n-1$. By exchangeability,

$$P[Y_{n+1}=1|S_n=k] = P[Y_{n+1}=1|S_{n-1}=k, Y_n=0]$$

and

$$P[Y_{n+1}=1|S_n=k+1] = P[Y_{n+1}=1|S_{n-1}=k, Y_n=1].$$

Now apply the previous case to the process $(Y_n, Y_{n+1}, \ldots)$ given $S_{n-1} = k$. □

From now on, assume that Y, in addition to being exchangeable, is an urn process with initial composition (r,b) and urn function f with domain $D = \{(r+k)/(m+n)\colon k = 0,1,\ldots, n;\ n = 0,1,\ldots\}$ where $m = r + b$.

<u>Lemma 2</u>. The function f is strictly increasing on D.

<u>Proof</u>: By definition of the urn function,

$$f\big((r+k)/(m+n)\big) = P[Y_{n+1}=1|S_n=k].$$

Now use Lemma 1. □

Let $i = r + k$ and $l = m + n$ where $0 \leq k \leq n$. Thus i/l is an element of D.

Notice that $i/(l + 1)$ and $(i + 1)/(l + 1)$ also belong to D.

Lemma 3. $f(i/l) - f\big(i/(l+1)\big) = f(i/l)\big[f\big((i+1)/(l+1)\big) - f\big(i/(l+1)\big)\big]$.

Proof. By the law of total probability and the definition of f,

$$\begin{aligned} P[Y_{n+2}=1|S_n=i] &= P[Y_{n+2}=1|S_n=i,Y_{n+1}=1]P[Y_{n+1}=1|S_n=i] \\ &\quad + P[Y_{n+2}=1|S_n=i,Y_{n+1}=0]P[Y_{n+1}=0|S_n=i] \\ &= f\big((i+1)/(l+1)\big)f(i/l) + f\big(i/(l+1)\big)\big(1-f(i/l)\big). \end{aligned}$$

By exchangeability and the definition of f,

$$P[Y_{n+2}=1|S_n=i] = P[Y_{n+1}=1|S_n=i] = f(i/l). \qquad \square$$

Lemmas 2 and 3 will be used to see that f has a continuous extension to all of the unit interval. Let $0 < \alpha < 1$. Because f is increasing, it has a left limit

$$f(\alpha-) = \sup\{f(x)\colon x < \alpha,\ x \in D\}$$

and a right limit

$$f(\alpha+) = \inf\{f(x)\colon x > \alpha,\ x \in D\}.$$

Lemma 4. For $0 < \alpha < 1$, $f(\alpha-) = f(\alpha+)$.

Proof: Choose sequences $\{i_m\}$ and $\{l_m\}$ such that

(a) $i_m/l_m \in D$ and $i_m/(l_m+1) < \alpha < i_m/l_m$ for all m,

(b) $l_m \to \infty$ as $m \to \infty$.

It follows that $i_m/(l_m+1)$ and i_m/l_m converge to α as $m \to \infty$.

By Lemma 3,

$$f(i_m/l_m) - f\big(i_m/(l_m+1)\big) = f(i_m/l_m)\big[f\big((i_m+1)/(l_m+1)\big) - f\big(i_m/(l_m+1)\big)\big].$$

Let $m \to \infty$ and examine both sides of this expression. The left-hand-side converges to $\Delta f(\alpha) \equiv f(\alpha+) - f(\alpha-)$. The right-hand-side converges to $f(\alpha+)\Delta f(\alpha)$. Now $0 < f(\alpha+) < 1$ by Lemma 2. So we conclude that $\Delta f(\alpha) = 0$. □

It follows from Lemmas 2 and 4 that f has a continuous extension to [0,1]. Because D is dense in [0,1], the extension is unique and there is no harm is denoting the extension by the same symbol 'f'.

Let X_n be the proportion of red balls in the urn at stage n; that is, $X_n = (r+S_n)/(m+n)$. As was mentioned in the introduction, the variables $Y_1, Y_2, \ldots$ are independent Bernoulli (θ) given $\Theta = \theta$. So the strong law of large numbers implies that S_n/n converges to Θ almost surely. Obviously, X_n converges to Θ almost surely also.

Recall that Θ has distribution Q and let S be the support of Q.

Lemma 5. For $x \in S$, $f(x) = x$.

Proof: By Corollary 3.1 of Hill, Lane, and Sudderth (1980), $f(\Theta) = \Theta$ almost surely. Thus the closed set $\{x: f(x) = x\}$ must contain S. □

To complete the proof of the theorem, let x be an element of D. It suffices to show $f(x) = x$ or, by Lemma 5, that x belongs to S.

Choose sequences of positive integers $\{k_j\}$ and $\{n_j\}$ such that $n_j \to \infty$ as $j \to \infty$ and $(r+k_j)/(m+n_j) = x$ for every j. Then, for each j, the event

$$A_j = [(r+S_{n_j})/(m+n_j) = x] = [S_{n_j} = k_j]$$

has positive probability by our assumption that $Q\{0,1\} < 1$. Condition on A_j and use the definition and continuity of the urn function to see that

$$(1) \quad E(\Theta|A_j) = P[Y_{n_j+1} = 1|A_j] = f(x)$$

and

$$\begin{aligned}(2) \quad \mathrm{Var}(\Theta|A_j) &= P[Y_{n_j+1} = Y_{n_j+2} = 1|A_j] - \big(P[Y_{n_j+1} = 1|A_j]\big)^2 \\ &= f(x)f\big((r+k_j+1)/(m+n_j+1)\big) - f(x)^2 \\ &\to 0 \quad \text{as} \quad j \to \infty.\end{aligned}$$

Let Q_j be the conditional distribution of Θ given A_j. By (1) and (2), Q_j converges to a distribution δ concentrated on the single point $f(x)$. Now each Q_j gives probability 1 to the closed set S because Q does. Hence, the limit measure δ also gives mass 1 to S. Thus $f(x) \in S$, and so by Lemma 5, $f\big(f(x)\big) = f(x)$.

Suppose $f(x) \neq x$. Say, $f(x) > x$. The strict monotonicity of f (Lemma 2) would give the contradiction $f(f(x)) > f(x)$.

The proof of the theorem is complete.

3. More than two colors.

Suppose $Y = (Y_1, Y_2, \ldots)$ is a sequence of variables with possible values in $\{1, 2, \ldots, k\}$. The process Y is a generalized k-color urn scheme if, for $n = 1, 2, \ldots$ and $1 \leq j \leq n$, the conditional probability $P[Y_{n+1}=j \mid Y_1=c_1, \ldots, Y_n=c_n]$ that the next ball is of color j depends only on the proportions of balls of each color at stage n. That is,

$$P[Y_{n+1}=j \mid Y_1=c_1, \ldots, Y_n=c_n] = f_j\left(\frac{r_1+n_1}{m+n}, \frac{r_2+n_2}{m+n}, \ldots, \frac{r_k+n_k}{m+n}\right)$$

where r_i is the number of balls of color i in the urn initially, n_i is the number of c_j's equal to i, and $m = \Sigma r_i$. The urn functions $f_1, \ldots, f_k$ have domains in the simplex $S_k = \{x = (x_1, \ldots, x_k): \Sigma x_i = 1, x_i \geq 0, i = 1, \ldots, k\}$ and are constrained to sum to 1.

If $f_j(x) = x_j$ for $1 \leq j \leq k$, then Y corresponds to a k-color Polya urn scheme in which at each stage a ball is drawn at random and replaced by two of the same color. In particular, Y is exchangeable and has de Finetti measure Q which is Dirichlet $(r_1, \ldots, r_k)$. (Note: For a k-valued exchangeable sequence Y, the de Finetti measure Q lives on the simplex S_k, and the variables $Y_1, Y_2, \ldots$ are independent, multinomial (θ) given $\Theta = \theta$ where Θ has distribution Q.) A natural conjecture, in the light of our theorem for 2-color processes, is that the de Finetti measure for a nondeterministic k-color exchangeable urn process is either a Dirichlet or a point mass. The following example shows that this conjecture is false. We do not know the correct characterization.

Example. Consider an urn which initially contains 1 red, 1 black, and 1 green ball. At each stage, a ball is selected at random. If it is red, it is replaced and another red ball is added. If the ball is black or green, it is replaced and another black or green is added depending on whether a fair coin falls heads or tails. This scheme corresponds to the urn functions

$$f_1(x) = x_1, \; f_2(x) = f_3(x) = \frac{x_2 + x_3}{2} .$$

The resulting urn process Y is easily seen to be exchangeable with de Finetti measure Q specified as follows: Q is a measure on $S_3 = \{\theta = (\theta_1, \theta_2, \theta_3)$: $\Sigma\theta_i = 1, \; \theta_i \geq 0 \quad i = 1,2,3\}$; under Q, θ_1 is beta (1,2) and, given θ_1, $\theta_2 = \theta_3 = \frac{1 - \theta_1}{2}$.

References

1. Arthur, B., Ermoliev, Yu.M., and Kaniovski, Yu.M. (1983). A generalized urn problem and its applications. Kibernetika 19 49-57 (in Russian). Translated in Cybernetics 19 61-71.

2. de Finetti, Bruno (1931). Funzione caratteristica di un fenomeno aleatorio. Atti Accad. Naz. Lincei. memo. Cl. Sci. Fis. Mat. nat. Sez. 6, 4 251-299.

3. de Finetti, Bruno (1937). La prevision: ses lois logigues, ses sources subjectives. Ann. Inst. H. Poincare 7 1-68. Translated in Studies in Subjective Probability (1968). ed. Kyburg, Smokler. Wiley, New York.

4. de Finetti, Bruno (1975). Theory of Probability, v. 2. Wiley, New York.

5. Freedman, David A. (1965). Bernard Friedman's urn.. Ann. Math. Statist. 36 956-970.

6. Hill, Bruce M., Lane, David, and Sudderth, William (1980). A strong law for some generalized urn processes. Ann. Prob. 8 214-226.

7. Johnson, Norman, and Kotz, Samuel (1977). Urn Models and Their Applications. Wiley, N.Y.

8. Polya, G. (1931). Sur quelques points de la theorie des probabilities. Ann. Inst. H. Poincare 1 117-161.